2012-01

Book Review: Analysis of Questionnaire Data with R

Fricker, Ronald D. Jr.

Fricker, R.D., Jr. (2012). Book Review: Analysis of Questionnaire Data with R, Journal of Statistical Software, 46, book review 1, January, 1-3.

Journal of Statistical Software

January 2012, Volume 46, Book Review 1. http://www.jstatsoft.org/

Reviewer: Ronald D. Fricker, Jr.
Naval Postgraduate School

Analysis of Questionnaire Data with R

Bruno Falissard
Chapman & Hall/CRC, Boca Raton, FL, 2012.
ISBN 978-1-4398-1766-7. 269 pp. USD 89.95 (H).
http://bruno.falissard.pagesperso-orange.fr/AQR/

Paraphrasing its preface, *Analysis of Questionnaire Data with R* is intended to help the non-statistician practitioner analyze questionnaire data using R. While laudable and desirable, it is not clear to me the book achieves this goal, at least for a general readership. I say this because, at least in my opinion, for adequate comprehension of the material the reader requires at least a working knowledge of R, sampling, and general survey methodology. More on this point later; first a description of the book.

Analysis of Questionnaire Data with R is divided into nine chapters that generally proceed from the very basics of analysis through some sophisticated methods. Throughout, the text uses data from a 2002 survey about mental disorder in French prisons to illustrate the methods. It liberally reproduces the necessary R code and output, most of which is available at a companion website. At 269 pages, the book is nicely compact, well organized, and for the reader who is already familiar with R, sampling, and survey methodology, it is quite easy to jump from section to section and read through them quickly.

The first chapter, in seven pages, defines what a questionnaire is, lays out five principles for analysis, and introduces R in a section entitled "If You Are a Complete R Beginner." Clearly much is abridged in these three topics – each of which could fill one or more chapters on their own – to fit into such a brief space. For example, the section on R provides the absolute minimum for a PC-based R user to be able to start up R and read in the example data set so that he or she can then copy and run the R commands in the subsequent chapters. I specified a PC-based R user because the illustrations show the "RGui" interface without any mention that the Apple and Unix/Linux interfaces are different. For sophisticated computer users, this chapter is probably sufficient to get started, but for others it is probably not. And, unfortunately, the author does not provide any pointers to good references for those readers who want or need additional background information (though Chapter 9 and an appendix, described below, do provide a bit more information).

Proceeding on through the book, Chapter 2 is focused on univariate descriptive statistics, both numerical and graphical. Chapter 3 discusses methods for quantifying and describing

relationships between variables, including correlation, relative risk, the odds ratio, and scatterplots, along with clustering and principle components. Chapter 4 describes various confidence intervals and hypothesis tests, as well as a brief discussion of sample size calculations. It is not until Chapter 4 that the text mentions complex sampling, and it does so only by briefly explaining how to use Thomas Lumley's **survey** package for the specific confidence interval and hypothesis testing calculations required for the French prison example's clustered sample design.

Chapter 5 discusses linear, logistic (binary, categorical, ordinal), and log-linear modeling, both under the assumption of simple random sampling and using the **survey** package (again for the prison example's clustered sample design). Chapter 6 delves a bit more deeply into issues related to modeling survey data, including coding variables, how to test categorical variables coded into multiple indicators, interaction terms, missing data, the bootstrap, and hierarchical modeling. In terms of missing data, a topic usually quite important in the analysis of survey data, after a somewhat extended discussion of methods for assessing the level of missingness for each variable, the text only briefly demonstrates how to use the `impute` and `mice` functions for imputation. Chapter 7 describes methods for deriving and validating a composite score derived from multiple survey questions, including factor analysis and Cronbach's alpha, and Chapter 8 introduces structural equation modeling. The text concludes with Chapter 9, which is focused on the basics of data manipulation in R, an appendix entitled "The Analysis of Questionnaire Data Using R: Memory Card," a list of references, and an index.

As I previously mentioned, *Analysis of Questionnaire Data with R* assumes a working knowledge of R, sampling, and general survey methodology. Indeed, while there are 4.5 pages of references, the book never mentions any of the standard texts on survey methodology, such as Groves, Fowler, Couper, Lepkowski, Singer, and Tourangeau (2009), Fowler (1995), and Dillman, Smyth, and Christian (2009). Nor does it mention any of the standard texts on sampling, such as Cochran (1977), Kish (1995), and Lohr (1999). And, even though it uses Lumley's R **survey** package, it does not reference his book (Lumley 2010), which provides much more detail on the R **survey** package.

The book is also narrowly focused on just the one French prison survey data set. While I appreciate the author's point that demonstrating the various analytical methods on actual data is helpful and sometimes illuminating, I would have liked more variety in the types of survey data presented and analyzed.

Based on the foregoing, in my opinion the book is not likely to be useful for introductory audiences. Nor is it particularly suitable as a text or a primer on survey analysis, unless it is used for an advanced course for which the students have already learned all the foundational concepts.

Readers to whom this book may be of interest are those possessing much of the requisite survey knowledge, but perhaps having only done survey analysis in another statistical package (e.g., SAS or Stata), and who are interesting in using R for a future analysis. Such readers would likely find the text insufficient for mastering R, but illuminating in how R can be used for survey analysis. Similarly, statisticians and other practitioners familiar with R, but perhaps not with survey analysis, may find the book useful for understanding how another obviously very capable researcher conducted a particular survey analysis with R. Indeed, I have found myself already referring to portions of the text as I consider various survey analyses, and I have recommended at least portions of it to students and colleagues.

In summary, this is an interesting and well-written book that will be of interest to particular, specialized audiences.

References

Cochran WG (1977). *Sampling Techniques*. 3rd edition. Wiley.

Dillman DA, Smyth JD, Christian LM (2009). *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 3rd edition. John Wiley & Sons.

Fowler Jr FJ (1995). *Improving Survey Questions: Design and Evaluation*. Sage Publications.

Groves RM, Fowler Jr FJ, Couper MP, Lepkowski JM, Singer E, Tourangeau R (2009). *Survey Methodology*. 2nd edition. John Wiley & Sons.

Kish L (1995). *Survey Sampling*. John Wiley & Sons.

Lohr SL (1999). *Sampling: Design and Analysis*. 2nd edition. Duxbury Press.

Lumley T (2010). *Complex Surveys: A Guide to Analysis Using R*. John Wiley & Sons.

Reviewer:

Ronald D. Fricker, Jr.
Naval Postgraduate School
Operations Research Department
Monterey, CA 93943, United States of America
E-mail: rdfricker@nps.edu
URL: http://faculty.nps.edu/rdfricke/

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 46, Book Review 1 *Published:* 2012-01-09
January 2012